2

New Parallel Models for Face Recognition

Heng Fui Liau, Kah Phooi Seng, Li-Minn Ang and Siew Wen Chin
University of Nottingham Malaysia Campus
Malaysia

1. Introduction

Face recognition has gained much attention in the last two decades due to increasing demand in security and law enforcement applications. Face recognition methods can be divided into two major categories, appearance-based method and feature-based method. Appearance-based method is more popular and achieved great success.

Appearance-based method uses the holistic features of a 2-D image. Generally face images are captured in very high dimensionality, normally is more than 1000 pixels. It is very difficult to perform face recognition based on original face image without reducing the dimensionality by extracting the important features. Kirby and Sirovich (Kirby & Sirovich, 1990) first used principal component analysis (PCA) to extract the features from face image and used them to represent human face image. PCA seeks for a set of projection vectors which project the image data into a subspace based on the variation in energy. In 1991, Turk and Pentland (Turk & Pentland, 1991) introduced the well-known eigenface method. Eigenface method incorporates PCA and showed promising results. Another well-known method is Fisherface (Belhumeur, 1997). Fisherface incorporates linear discriminant analysis (LDA) to extract the most discriminant features and to reduce the dimensionality. In general, LDA-based methods outperform PCA-based methods because LDA optimizes the low- dimensional representation of face images with the focus on the most discriminant features extraction. LDA seeks for a set of projection vectors which form the maximum between-class scatter and minimum within-class scatter matrix simultaneously (Chen et al, 2000).

More recently, frequency domain analysis methods such as discrete Fourier transform (DFT), discrete wavelet transform (DWT) and discrete cosine transform (DCT) have been widely adopted in face recognition. Frequency domain analysis methods transform the image signals from spatial domain to frequency domain and analyze the features in frequency domain. Only limited low-frequency components which contain high energy are selected to represent the image. Unlike PCA and LDA, frequency domain analysis methods are data independent. They analyze image independently and do not require training images. Furthermore, fast algorithms are available for the ease of implementation and have high computation efficiency.

In this chapter, new parallel models for face recognition are presented. Feature fusion is one of the easy and effective ways to improve the performance. Feature fusion method is performed by integrating multiple feature sets at different levels. However, feature fusion method does not guarantee better result. One major issue is feature selection. Feature

Source: Recent Advances in Face Recognition, Book edited by: Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett,
ISBN 978-953-7619-34-3, pp. 236, December 2008, I-Tech, Vienna, Austria

selection plays a very important role to avoid overlapping features and information redundancy. We propose a new parallel model for face recognition utilizing information from frequency and spatial domains. Both features are processed in parallel way. It is well-known that image can be analyzed in spatial and frequency domains. Both domains describe the image in very different ways. The frequency domain features are extracted using DCT, DFT and DWT methods respectively. By utilizing these two very different features, a better performance is guaranteed.
Feature fusion method suffers from the problem of high dimensionality because of the combined features. It may also contain redundant and noisy data. To solve this problem, LDA is applied on the features from frequency and spatial domains to reduce the dimensionality and extract the most discriminant information. However, LDA has a big drawback. If the number of samples is smaller than the dimensionality of the samples, the sample scatter matrix may become singular or close to singular, leading to computation difficulty. This problem is called small sample size (SSS) problem. Several variants of LDA have been developed to counter SSS problem such as, Liu LDA (Liu et al, 1992), Chen LDA (Chen et al, 2000), D-LDA (Hu & Yang, 2001) and modified Chen LDA. These modified LDA techniques will be presented and discussed. Different variants of our parallel model face recognition with different frequency domain transformation techniques and variants of LDA algorithms are proposed. The strategy of integrating the multiple features is also discussed. A weighting function is proposed to ensure the features from spatial and frequency domains contribute equal weight in the matching score level.
ORL and FERET face databases were chosen to evaluate the performance of our system. The results showed that our system outperformed most of the conventional methods.

2. Frequency domain analysis methods

Frequency domain analysis method has been widely used in modern image processing. In this section, DFT, DCT and DWT are presented.

2.1 Discrete fourier transform

Fourier Transform is a classical frequency domain analytical method. For an $1 \times N$ input signal, $f(n)$. DFT is defined as

$$F(k) = \int_{n=1}^{N} f(n) e^{-j2\pi(k-1)(\frac{n-1}{N})}\, dt \qquad 1 \leq k \leq N \tag{1}$$

The 2D face image is first converted to 1D vector, $f(n)$ by cascading each column together and transforming them into frequency domain. Only low frequency coefficients are selected because most of the signal's energy is located in the low frequency band. In this chapter, 300 coefficients (from k=1 until k=300) are selected. As a matter of fact, human visual system is more sensitive to variation in the low-frequency band [10].

2.2 Discrete cosine transform

DCT possesses some fine properties, such as de-correlation, energy compaction, separability, symmetry and orthogonality. According to the JPEG image compression standard, the image is first divided into 8×8 blocks for the purpose of computation efficiency. Then, two dimensional DCT (2D-DCT) is applied independently on each block.

The DCT coefficients are scanned in a zigzag manner starting from the top left corner of each block as shown in Fig. 1 because DCT coefficients with large magnitude are mainly located at the upper left corner. The first coefficient is called DC-coefficient. The remaining coefficients are referred to as AC coefficients. The frequency of the coefficients increases from left to right and from top to bottom. The DCT coefficients at the most upper-left corner of each 8×8 block are selected and merged to a 1D vector. For an $N \times N$ image, the 2D DCT is defined as

$$C(u,v) = \propto(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\frac{\pi(2x+1)u}{2N}\cos\frac{\pi(2y+1)}{2N} \tag{2}$$

For υ , v = 0,1,2,…N-1 and $\alpha(u)$ and $\alpha(v)$ are defined as follow: $\alpha(u)=\sqrt{\frac{2}{N}}$ for υ=0 , and

$\alpha(v)=\sqrt{\frac{2}{N}}$ for $v \neq 0$.

Based on (Lay and Guan, 1999) and (Tjahyadi et al, 2007) works, *DC* and AC01, AC10, AC11 which are located at the top-left corner of the block are selected because they give the best result. LDA is further applied to the selected coefficient to extract the most discriminant features for the ease of computation and storage.

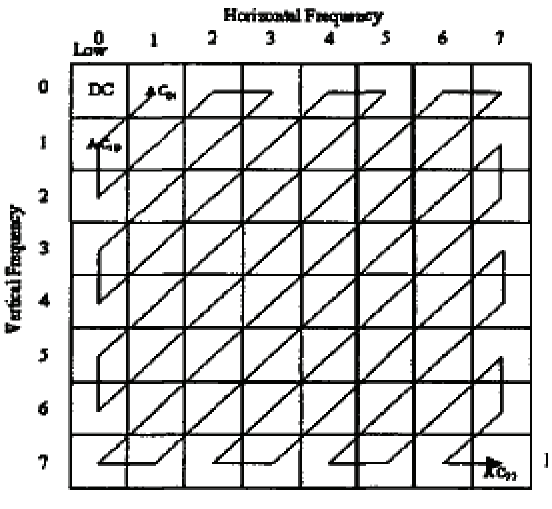

Fig. 1. The zigzag scanning pattern in DCT block

2.3 Discrete wavelet transform

DWT has been widely employed for noise reduction and compression in modern image processing. DWT operates by performing convolution on a target signal with wavelet kernel. There are several well-known wavelets such as coif (3), Haar and etc. DWT decomposes a signal into a sum of shifted and scaled wavelets. The continuous wavelet transform between a signal $f(t)$ and a wavelet $\varphi(n)$ is defined as

$$c(a,b) = \frac{1}{\sqrt{a}} \int f(t)\varphi\left(\frac{t-b}{a}\right)dt \tag{3}$$

where a is the scale and t is the time, and b is the shift. For DWT, the scale, a is restricted to powers of 2 and the position, b, is restricted to the integers multiples of the scales. DWT is defined as

$$c_{j,k} = \int_{-\infty}^{\infty} x(t)\varphi_{j,k}dt \tag{4}$$

where j and k are integers and $\varphi_{j,k}$ are orthogonal baby wavelets defined as

$$\varphi_{j,k} = 2^{\frac{j}{2}}\varphi(2^j t - k) \tag{5}$$

Baby wavelets $\varphi_{j,k}$ have an associated baby scaling function defined as

$$\emptyset_{j,k}(t) = 2^{\frac{j}{2}}\emptyset(2^j t - k) \tag{6}$$

The scaling function can be expressed in terms of low-pass filter coefficients $h_0(n)$ as shown below:

$$\emptyset(t) = \sum_n h_1(n)\sqrt{2}\emptyset(2t-n) \tag{7}$$

The wavelet function can be expressed in term of high-pass filter coefficients $h_1(n)$ as below

$$\varphi(t) = \sum_n h_1(n)\sqrt{2}\emptyset(2t-n) \tag{8}$$

Hence, the signal $f(t)$ can be rewritten as below:

$$f(t) = \sum_k cA_1(k)\emptyset_{j-1,k}(t) + \sum_k cD_1(k)\varphi_{j-1,k}(t) \tag{9}$$

Where $cA_1(k)$ and $cD_1(k)$ represent the approximation coefficients and detail coefficients level 1 respectively. Similarly, the approximation and detail coefficient can be expressed in term of low-pass filter coefficients, $h_0(n)$ and high-pass filter coefficients, $h_1(n)$.

$$cA_1(k) = \sum_n cA_0(n)h_0(n-2k) \tag{10}$$

$$cd_1(k) = \sum_n cA_0(n)h_1(n-2k) \tag{11}$$

2-D DWT is implemented by first computing the one-dimensional DWT along the rows and then columns of the image (Meada et al, 2005) as shown in Fig. 2. Features in LL sub-band are corresponding to low-frequency coefficients along the rows and columns and all of them are selected to represent the face image.

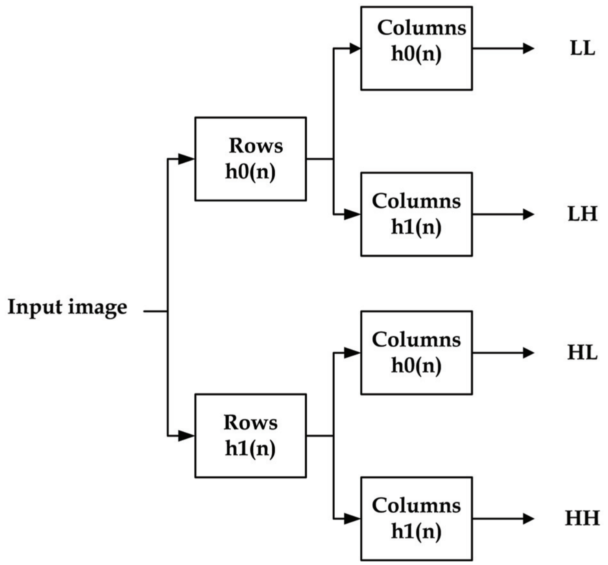

Fig. 2. Two-dimensional discrete cosine transform

3. Linear discriminant analysis

As mentioned in the previous section, feature fusion method suffers from the problem of high dimensionality. Our proposed method incorporates LDA to reduce the dimensionality of the features from frequency and spatial domains. Conventional LDA seeks for a set of projection vectors, W which form the maximum between-class scatter, S_b and minimum within-class scatter matrix, S_w simultaneously (Chen et al, 2000). The function of W is given in Eq. (14).

$$S_w = \sum_{j=1}^{K}\sum_{i=1}^{M}(x_i^j - m_j)(x_i^j - m_j)^T \tag{12}$$

$$S_B = \sum_{j=1}^{K}(m_j - m)(m_j - m)^T \tag{13}$$

$$W = \arg\max \frac{W^t S_b W}{W^t S_w W} \tag{14}$$

For a database which contains K classes and each class has M samples, each sample is represented by n-dimensional vector. The rank of S_w is defined as in Eq. (12). LDA algorithm has a big drawback which is SSS problem. Liu et al, Yang et al and Chen et al proposed different approaches to handle the SSS problem.

If the rank of $S_w \neq n$, then S_w is singular. Liu et al modified the traditional LDA algorithm by replacing S_w in Eq. (14) with total scatter matrix, S_t. S_t is the sum of within-class scatter matrix and between-class scatter matrix. The new projection vector set is defined as in Eq.

(17). The rank of S_t is defined as in Eq. (16) as shown in (Chen et al, 2000).If $S_t \neq n$, S_t is non-singular. Under this circumstance, the LDA criteria will be fulfilled if $W^tS_wW=0$ and $W^tS_bW\neq 0$. Although KM-1> $K(M$-1) , this does not guarantee that S_t is always not equal to n.

$$rank(S_w) = \min(n, K \times (M-1)) \tag{15}$$

$$rank(S_t) = \min(n, KM-1) \tag{16}$$

$$W = \arg\max \frac{W^tS_bW}{W^tS_wW + W^tS_bW} \tag{17}$$

Yang et al proposed a solution called D-LDA to solve the small sample size problem. Unlike conventional LDA, D-LDA starts by diagonalizing the between-class scatter matrix S_b. All of the eigenvectors of which the corresponding eigenvalues are equal to zero or close to zero are discarded because they do not carry any discriminative power (Hu and Yang, 2001). The remaining eigenvectors and the corresponding eigenvalues are chosen to form D_b and V_b respectively. Then, the within-class scatter matrix S_w is transformed to S_{ww}. S_{ww} is defined as below:

$$S_{ww} = \left(D_b V_b^{-\frac{1}{2}}\right) S_w \left(D_b V_b^{-\frac{1}{2}}\right) \tag{18}$$

The projection vector that can satisfy the objective of an LDA process is the one that can maximize the between-class scatter matrix. Only the smallest eigenvalues and the corresponding eigenvalues are chosen to form V_W and D_W respectively. The most discriminant vector set for D-LDA is given by

$$U = D_w^{-\frac{1}{2}} (D_b V_b)^T (D_w)^T \tag{19}$$

Chen LDA used a different approach to counter the problem. Chen LDA starts by calculating the projection vector in the null space of the S_w. This is done by performing singular value decomposition on S_w. Then a set of eigenvectors, of which corresponding eigenvalues are equal to zero, are chosen to form the projection vector. The projection vector set projects S_b to another subspace and the new S_b is $\widetilde{S_b}$. Singular value decomposition is performed on $\widetilde{S_b}$. A set of projection vector, in which corresponding eigenvalues are the largest are chosen. Now, there are two set of eigenvectors. A set of eigenvectors is derived from the null space of S_w. Another set of eigenvectors is derived from S_b, in which the corresponding eigenvalues are the largest. With both set of eigenvectors, the objective of LDA is fulfilled. Chen LDA is summarized as below:

Step 1, Perform the singular value decomposition of S_w. Choose a set of eigenvectors, in which the corresponding eigenvalues are zero to form Q.

Step 2, Compute S_{bb}, where $S_{bb}=QQ^tS_b(QQ^t)^t$. S_b is the between-class scatter matrix.

Step 3, Perform the singular value decomposition of S_{bb}. Choose a set of eigenvectors, in which the corresponding eigenvalues are the largest, to form U. U is the most discriminant vector set for LDA.

In this chapter, Chen LDA algorithm is modified. Instead of only choosing the eigenvectors which the corresponding eigenvalues are equal to zero in the *step 1*, we further includes those eigenvectors which the corresponding eigenvalues are close to zero. We deduced that the most discriminant features are not only located in null space of S_w but also eigenvalues that close to zero. By selecting more eigenvectors, the most discriminant information in S_w is preserved.

4. Parallel models for face recognition

As mentioned in previous section, LDA is applied on the features extracted from frequency and spatial domains. There are two set of features. One carries the important information of the face image which is derived from the spatial domain and the other one from frequency domain. Both sets of feature describe the face images in very different way. Here, both feature sets are assumed to be equally important. In order to make both features from spatial and frequency domains give equal weight in total matching score, a weighting function is applied to the feature set from spatial domain. The weighting function is given in Eq. (20).

$$\omega = \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} f_i} \tag{20}$$

Given that S is the feature from spatial domain and f is the feature from frequency domain. The sizes of both features are 1×n. The weighting function is applied to the spatial domain features. The feature vectors from both domains are merged into 1-D vectors [f1,f2,…fn, ωs1, ωs2,…, ωsn].
In section 3, the problem of LDA had been discussed. Chen LDA, D-LDA and modified Chen LDA are capable to counter SSS problem. But Chen LDA and modified Chen LDA do not perform well when S_w is non-singular. Liu LDA cannot counter SSS problem when Eq. (16) equal to n. D-LDA can perform well regardless the condition of S_w because D-LDA starts calculating S_b instead of S_w. Our results in section 5 showed that Liu LDA and D-LDA are equally good when S_w is non-singular. Modified Chen LDA gave the best result when S_w is singular. Based on the simulation result in section 5, three variants of our parallel model face recognition system as shown in Figure 3 are developed. The selection of LDA algorithm is based on the choice of feature domain. The selected DCT features from DCT domain in ORL database in small and the corresponding S_w is non-singular. Hence, D-LDA is incorporated to extract the most discriminant features and to further reduce the dimensionality. D-LDA has advantage over Liu LDA in term of computation because D-LDA does not involve matrix inversion. For DWT and DFT, the feature sets are relatively large and S_w is singular. Modified Chen LDA is employed to extract the most discriminant features because it gave the best result when S_w is singular.

5. Simulation results

The Olivetti Research Laboratory (ORL) and FERET databases were chosen to evaluate the performance of our proposed system. ORL database contains 400 pictures from 40 persons, each person has 10 different images. For each person, 5 pictures are randomly chosen as the training images. The remaining 5 pictures serve as the test images. The similarity between

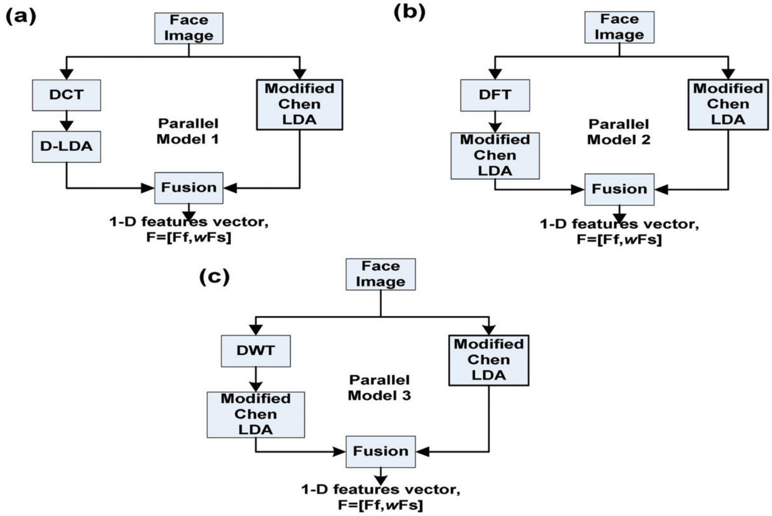

Fig. 3. Parallel models for face recognition

two images is measured using Euclidean Distance. Shorter distance implies higher similarity between two face images. *fb* probe set from FERET database was chosen to evaluate the proposed methods. The training set consists of 165 frontal images from 55 people. Each person has 3 different frontal images.

5.1 Spatial domain result

The dimensionality of the face image was 32×32. ORL database is chosen to evaluate the performance. According to Eq. (15) and Eq. (16), S_w and S_t are singular. Hence, Liu LDA cannot solve the problem. Chen LDA, modified Chen LDA and D-LDA are employed to extract the most discriminant information and further reduce the dimensionality of the feature set from spatial domain. PCA result is included for comparison purpose. The performance for each system is shown in Table 1.

Method	Recognition Rate (%)
PCA	89.5
Chen LDA	90.5
D-LDA	89.5
Modified Chen LDA	91.5

Table 1. Spatial domain result

As shown above, the modified Chen LDA gave the best result. We deduced that modified Chen LDA gave the best result because it preserved more discriminant information of S_w compared to Chen LDA. Hence, modified Chen LDA will be employed to extract the feature when the sample encounter SSS problem.

5.2 Frequency domain result

Since there were only 4 coefficients selected from each block, the total number of coefficients was 64. According (3) and (4), S_w and S_t are non-singular and LDA can be performed in DCT domain without difficulty. Liu LDA and D-LDA were employed to extract the most discriminant features. For DFT and DWT, the number of selected features that represent face image is 300 and 400 respectively. Therefore, Chen LDA, modified Chen LDA and D-LDA are incorporated to extract the most discriminant features.

From Table 2, it can be seen that Liu LDA and D-LDA gave equally good result in DCT domain which the sample does not suffer SSS problem. They achieved 94% recognition rate. For DFT and DWT which both S_w were singular, modified Chen LDA gave the best result. It scores 96.5% and 94% in DFT domain and DWT domain respectively. Among the frequency domain analysis method, DFT gave better result compared to others. DFT + modified Chen LDA gave the best result.

Method	Recognition rate (%)
DCT+ Liu LDA	94
DCT+D-LDA	94
DFT+Chen LDA	94
DFT+modified Chen LDA	96.5
DFT+ D-LDA	92
DWT+Chen LDA	91
DWT+modified Chen LDA	93.5
DWT+ D-LDA	90.5

Table 2. Frequency domain result

5.3 Parallel models for face recognition result

All parallel models outperformed most of the conventional methods as shown in Table 3. Parallel model 2 gave the best result. Both of them achieved 99% recognition rate in ORL database. Parallel model 2 outperformed parallel model 2 and 3 because the corresponding frequency domain features gave better result. Parallel model 2 and 3 only achieved 97.5% and 96.5% recognition rate respectively.

Method	Recognition rate (%)
Parallel Model 1	97.5
Parallel Model 2	99
Parallel Model 3	96.5
D-LDA (Hu and Yang, 2001)	94
DWT+SHMN (Amira et al, 2007)	97
FD-LDA (Lu et al, 2003)	96

Table 3. Comparison of recognition rate of other face recognition methods

The performances of the proposed parallel models are further evaluated using *fb* probe set of FERET database. Fig. 5. shows the recognition rate of the proposed methods under different number of features. Fig. 6. shows the cumulative matching score (CMS) curve of the proposed methods. Since there are 165 classes, the number of output features of LDA is

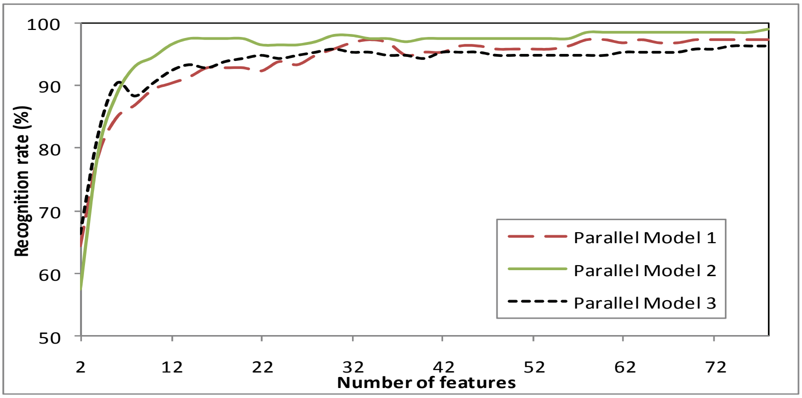

Fig. 4. ORL database result

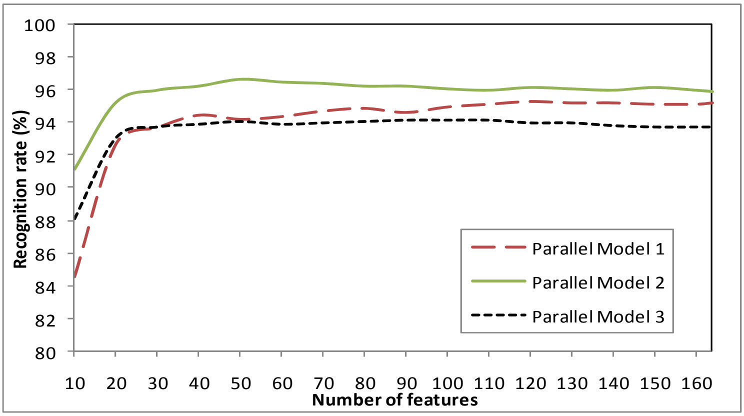

Fig. 5. FERET result

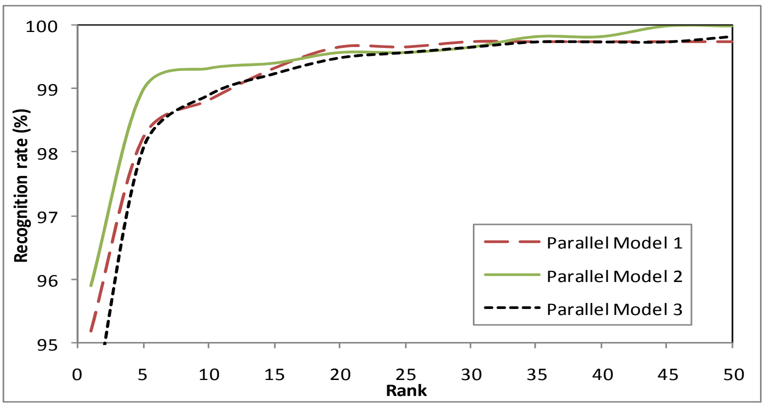

Fig. 6. CMS curve

164. Therefore, the number of selected coefficients from DCT domain is increased from 64 to 192 for parallel model 1.

Similar to ORL database's result, parallel model 2 gave the best result. It achieved 96.7% recognition rate when the number of features was 50. It also gave the best result in CMS. It achieved 100% recognition rate when the rank was 45 and above.

6. Conclusion

In this paper, a new parallel model for face recognition is proposed. There are three variants of parallel model which incorporate different variants of LDA. The proposed utilizing information form frequency and spatial domains. Both features are processed in parallel way. LDA is subsequently applied on the features to counter high dimensionality problem that encounter by feature fusion method. The high recognition rate that is achieved by the proposed methods shows that features of both domains contribute valuable information to the system. Parallel model 1 and 2 gave the best result. Parallel model 2 achieved 99% and 96.7% recognition rate in ORL and FERET database respectively.

7. References

Belhumeur, P.N.; Hespanha, J.P. & Kriegman, D.J. (1997) Eigenface vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Machine Intell*, vol.19, pp.711-720, May 1997.

Chen, L.F.; Mark Liao, H.Y.; Ko, M.T.; Lin, J.C. & Yu, G.J. (2000) A new LDA-based face recognition system which can solve the small space size problem, *Pattern Recognition*, vol.33, pp.1703-1726, 2000.

Kirby, M. & Sirovich, L. (1990) Application of the Karhunen-Loeve procedure of the characteristic of human faces, *IEEE Trans. Pattern Anal.Machine Intell*, vol.12,pp 103-108, Jan,1990.

Lay, J.A. & Guan, L. (1999) Image Retrieval based on energy histogram of the low frequency DCT coefficients, *IEEE International Conference on Acoustics Speech and Signal Processing*, 6:3009-3012, 1999.

Liu, K.; Cheng, Y. & Yang, J. (1992) A generalized optimal set of discriminant vectors, *Pattern Recognition* vol. 25, no. 7, pp. 731-739, 1992.

Lu, J.; Plataniotis, K.N. & Venetsanopoulos, A. N. (2003) Face Recognition Using LDA-based Algorithm", *IEEE trans.Neural Network*, vol.14, No 1, pp.195-199, January 2003.

Meada, M.; Sivakumar, S.C. & Phillips, W.J. (2005) Comparative performance of principal component analysis, Gabor wavelets and discrete wavelet transforms for face recognition", Can. J. Elect. Comput. Eng, vol. 30, No. 2, 2005.

Nicholl, P.; Amira, A.; Bouchaffra, D. & Perrott, R.H. (2007) Multiresolution Hybrid Approaches for Automated Face Recognition, *AHS*, 2007.

Tjahyadi, R.; Liu, W.; An, S. & Venkatesh, S. (2007) Face Recognition via the Overlapping Energy Histogram, *IJCAI*, pp.2891-2896, 2007.

Turk, M. & Pentland, A. (1991) Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Mar 1991.

Yu, Hu. & Yang, J. (2001) A Direct LDA algorithm for high-dimension data with application to face recognition, *Pattern Recognition*, vol.34, pp. 2067-2070, 2001.